VERIFICATION OF TRANSLATION

I, Eva Lanyi, a translator with Chillson Translating Service, 3530 Chas Drive, Hampstead, Maryland, 21074, hereby declare as follows:

That I am familiar with the French and English languages;

That I am capable of translating from French to English;

That the translation attached hereto is a true and accurate translation of the French patent titled, "PROCESS FOR IN VITRO CREATION OF RECOMBINANT POLYNUCLEOTIDE SEQUENCES BY ORIENTED LIGATION;"

That all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true;

And further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any registration resulting therefrom.

By Eva Lanyi

Executed this 15 day of Oct 2001.

Witness Anne Chillson

PROCESS FOR IN VITRO CREATION OF RECOMBINANT POLYNUCLEOTIDE SEQUENCES BY ORIENTED LIGATION

This invention relates to a method of obtaining, *in vitro*, recombinant polynucleotide sequences by oriented ligation. The invention aims in particular at creating, then selecting, polynucleotide sequences able to offer one or several advantageous properties compared to corresponding properties of reference sequences and thus able to provide an improved phenotype and/or to produce improved proteins.

Reference sequence is understood as a sequence having properties close to those being sought.

In vitro is understood as any event, reaction, or process that does not take place in a living organism.

Ligation is understood as a process that allows the creation of a phosphodiester bond between two nucleic acid fragments.

Oriented ligation is understood as any process of ligation that makes it possible to assemble nucleic acid molecules in a set order, notably by hybridization of said nucleic acid molecules on at least one nucleotide matrix.

Polynucleotide sequence is understood as any single- or double-stranded nucleic acid molecule.

Various techniques have been developed to promote *in vitro* recombination among different polynucleotide sequences, among them can be cited, in particular, DNA shuffling and StEP, both based on the use of PCR.

DNA shuffling comprises two steps, the random fragmentation of polynucleotide sequences by DNAase I and amplification by PCR, in which the previously created fragments act as initiators. At each hybridization step,

the change of matrix causes recombinations at regions having homologous sequences. StEP consists in mixing various polynucleotide sequences containing various mutations in the presence of a pair of initiators. This mixture is subjected to a PCR reaction in which the hybridization and polymerization steps are consolidated into a single, very brief step. These conditions make it possible to hybridize the initiators but reduce the polymerization speed so that the fragments that are partially synthesized hybridize randomly on the polynucleotide sequences carrying the various mutations, thus making the recombination possible. In each of these two methods, the polymerization step is indispensable to the recombination process. Thus, depending on the polymerases selected, this polymerization step can produce undesired supplementary mutations. Further, after a certain number of cycles, DNA shuffling and StEP are based on the principle of hybridization of a "mega-initiator" on a matrix, which probably causes difficulties in implementation for polynucleotide sequences whose size is greater than 1.5 Kpb. Finally, these two techniques do not allow the rate of recombinations to be controlled, since the latter occur randomly during the successive stages of polymerization.

This invention aims specifically at reducing the above drawbacks by offering a simple preparation method for at least one recombinant polynucleotide sequence by using a process of oriented ligation of fragments obtained from a bank of polynucleotide sequences.

Bank of polynucleotide sequences is understood as a group of polynucleotide sequences containing at least two heterologous polynucleotide sequences.

In one embodiment of the invention, this object is achieved thanks to a process comprising the following steps:

a) fragmentation of a bank of polynucleotide sequences,

b) denaturation of the fragments thus obtained,

c) hybridization of fragments obtained in step (b) with one or several assembly matrix (matrices),

d) oriented ligation of said fragments to obtain at least one recombinant polynucleotide sequence.

When the assembly matrix is double-stranded, it is first denatured before step (c) such as, for example, during step (b).

The process of the invention makes it possible randomly to recombine various fragments within a polynucleotide sequence during steps (b), (c) and (d). This process thus reproduces, *in vitro*, the recombination phenomena that can occur *in vivo* by promoting them. The process of the invention is thus most particularly useful for recombining polynucleotide sequences among themselves to be able to create new polynucleotide sequences.

These recombinant polynucleotide sequences are able to offer advantageous properties compared to corresponding properties of references sequences and are thus able to provide an improved phenotype and/or to produce improved proteins.

This object is achieved thanks to a process comprising the following steps:

a) fragmentation of a bank of polynucleotide sequences,

b) denaturation of the fragments,

c) hybridization of fragments obtained in step (b) with one or several assembly matrix (matrices),

d) oriented ligation of said fragments to obtain at least one recombinant polynucleotide sequence,

e) selection of recombinant polynucleotide sequences offering

advantageous properties compared to corresponding properties of one or several references sequences.

The assembly matrix (matrices) can be single- or double-stranded. If one of these matrices is double-stranded, it is first denatured to be added in a single-stranded form to step (c), for example, during step (b).

The process of the invention can comprise, at the end of step (e), the repetition of steps (a), (b), (c), and (d). In this case, the bank of polynucleotide sequences contains at least one recombinant polynucleotide sequence that was selected in (e).

The process of the invention can also comprise, at the end of step (d) and before step (e), the repetition of steps (b), (c), and (d), or of steps (a), (b), (c), and (d).

This last embodiment is particularly useful if, at the end of step (e), all the fragments are not ligated. In that case, the process of the invention also comprises, at the end of step (d) and before step (e), one or several of the following reaction cycles:

- denaturation of ligated and non-ligated fragments from step (d), optionally in the presence of one or several assembly matrix (matrices),
- hybridization of said fragments with one or several assembly matrix (matrices) if the latter is (are) not present during the denaturation,
- ligation of said fragments.

These reactions of denaturation, hybridization and ligation are equivalent to steps (b), (c), and (d) but are performed not with fragments of step (a) but with ligated and nonligated fragments from step (d).

According to a particular embodiment of the process, the polynucleotide

sequences of the bank are single-stranded. The use of single-stranded DNA fragments is particularly adapted to the recombination of gene families for which a given single-stranded matrix or a mix of various single-stranded matrices will be hybridized with single-stranded fragments coming from a bank of homologous genes. Since there are no strictly complementary sequences among them in the fragment population, the hybridization will not be biased toward wild sequences among the fragments or among matrix fragments. The hybridization temperature can thus be adjusted depending on how homologous the sequences are, thus producing recombinant molecules with the greatest possible degree of recombination. Banks of recombinant molecules are thus created, with higher value in terms of diversity, considerably increasing the chances of finding the good mutant at the end of the minimum recombination cycles.

To obtain single-stranded DNA molecules, a Bluescript phagemide or a vector of the family of filamentous phages such as M13mp18 can be used. Another method consists in creating double-stranded molecules by PCR by using an initiator phosphorylated at 5' and the other non-phosphorylated. The digestion of the lambda phage by the exonuclease will destroy the strands of DNA phosphorylated at 5;, leaving the non-phosphorylated strands intact. Another method of creating single-stranded molecules consists in making an amplification, by asymmetric PCR, starting from a methylated DNA matrix.

Digestion by Dpn I will destroy the methylated strands, leaving intact the amplification products that will then be able to be purified after denaturation.

The process of the invention can further comprise one or several of the following steps:

- separation of recombinant polynucleotide sequences from the

assembly matrix (matrices)

before step (e),

- amplification of recombinant polynucleotide sequences before step (e),
- cloning of recombinant polynucleotide sequences optionally after separation of the recombinant strands from the matrix (matrices).

In one advantageous embodiment of the method, the ends of the fragments created at step (a) are such that there can be adjacent hybridization of these ends at least with one assembly matrix in step (c) and ligation of these fragments with one another at step (d). The polynucleotide sequences of the bank on which the process of the invention is performed must be such that the fragments obtained during the process have ends like those described above. These fragments can be obtained notably during step (a) or during step (d) by ligation of the fragments.

On advantageous embodiment of the process of the invention consists in simultaneously performing steps (c) and (d) as a so-called RLR reaction, in the English expression "recombining ligation reaction."

Besides the previously indicated advantages, the process of the invention is remarkable in that it promotes and accelerates the random *in vitro* recombination of polynucleotide sequences, these polynucleotide sequences being able to be genes. Gene is understood to be a fragment or a DNA sequence associated with a biological function. A gene can be obtained in different ways, among them chemical synthesis, synthesis by polymerization or extraction of said gene from a nucleic acid source.

The *in vitro* recombination of polynucleotide sequences from the initial bank by the process of the invention thus makes it possible to obtain a new

bank containing sequences having acquired one or several characteristics of the sequences of the preceding bank. The process of the invention thus constitutes a technique for *in vitro* evolution.

The process of the invention constitutes an alternative to recombinant PCR such as the use of the techniques of DNA shuffling or StEP, since it does not require the *in vitro* polymerization step to achieve recombination. On the contrary, the key step in the process of the invention is step (d) of ligation on an assembly matrix (or oriented ligation), which guarantees a very high degree of fidelity during the recombination events.

The process of the invention is remarkable in that it makes it possible considerably to increase the efficiency of reassembling the fragments to be ligated by using oriented ligation. In fact, in the case of a sequence cut into n fragments, there are very numerous possibilities of reassociation of the fragments using a conventional ligation process (without using a reassembly matrix that orients the ligation), among which only one form is useful. In the process according to the invention, the ligation is oriented by the assembly matrix, which makes it possible to obtain the only useful form directly.

The fragmentation of these polynucleotide sequences in step (a) can be performed in a controlled manner or in a random manner.

If fragmentation is performed in a controlled manner, the fragmentation makes it possible precisely to control the degree of recombination desired and the position of the recombination points. According to a preferred embodiment of the process of the invention, step (a) consists in subjecting the polynucleotide sequences of the bank to hydrolysis by the action of one or several restriction enzymes. Thus, in a particular embodiment of the process of the invention, the degree of recombination and the position of the

recombination points of the recombinant polynucleotide sequences are determined by the fragmentation of step (a).

Thus the greater the number of fragments produced per sequence, the greater the number of fragments necessary to recompose a sequence, which causes an increased rate of recombination. Further, the nature and the position of the ends of the fragments produced in the embodiment of the process of the invention can be known and controlled, which makes it possible to:

- precisely control the zones in which the recombination takes place, or
- to induce recombination among polynucleotide sequences, for example genes, if the ends of the fragments are created in zones that are homologous among these sequences, or in homologous zones between these sequences and the assembly matrices.

If the fragmentation is random, any enzymatic or mechanical means known to one skilled in the art and able randomly to cut the DNA can be used such as, for example, digestion by DNAase I or ultrasonication.

The process of the invention makes it possible considerably to increase the efficiency of reassembling the fragments to be ligated, it can thus be applied to the orientation of multimolecular ligation for flush ends. In this application, single- or double-stranded oligonucleotides that are exactly complementary to end 3' of a fragment and 5' of the adjacent fragment are used as the assembly matrix for steps (b) or (c), which makes possible the adjacent hybridization of these two ends on the same matrix after the denaturation step. Once hybridized, the ends of the fragments can be ligated among themselves so as to orient the ligation direction of the fragments at flush ends. The same approach can be envisaged for the orientation of the

ligation of fragments at cohesive ends.

A rather preferred embodiment of the process of the invention consists in adding enzymes to step (c) and/or to step (d) that are able to recognize and degrade and/or cut in a specific way the nonhybridized ends of fragments when the latter cover other hybridized fragments on the same matrix. A preferred example of this type of enzyme is the enzyme Flap endonuclease.

A particular embodiment of the process of the invention thus consists in using enzymes of the Flap endonuclease type when the fragments produced in step (a) can be covered during hybridization on the assembly matrix at step (c).

Thus, during the hybridization of DNA fragments on a matrix, these enzymes are characterized by the ability to recognize and to cut in a specific way the nonhybridized ends of these fragments when the latter cover other hybridized fragments on the same matrix.

When the fragments used during the process of the invention are double-stranded, a particular embodiment of the invention consists in using specific, single-stranded enzymes of the exonuclease type. These enzymes will be characterized by being able to recognize and to degrade in a specific way the single-stranded, nonhybridized ends of these fragments when the latter cover other hybridized fragments on the same matrix.

During step (c), hybridization, the use of this type of enzyme (notably Flap, or a single-stranded specific exonuclease) thus makes it possible to increase the number of adjacent ends able to be ligated in step (d), which is particularly significant if the fragments are obtained by random cutting, because these fragments have zones where they cover one another when they are hybridized on the assembly matrix.

In a particular embodiment of the process of the invention using an

active, preferably thermostable ligase at high temperature at step (d), the enzymes able to recognize and/or to cut in a specific way the nonhybridized ends of the fragments, added at step (c) and/or at step (d), will have the same characteristics of thermoresistance and high-temperature activity as said ligase.

The bank of polynucleotide sequences on which the process of the invention is performed can be produced by any method known to one skilled in the art, for example starting from a wild-type gene, by successively managed stages of mutagenesis, by "error prone" PCR (2), by random, chemical mutagenesis, by random mutagenesis *in vivo*, or by combining genes of close or distinct families within the same species or different species so as to make available in said bank a variety of polynucleotide sequences.

Among these techniques, the invention envisions most particularly a process in which the bank of polynucleotide sequences is obtained by a chain polymerization reaction performed under conditions that make it possible to create random, localized mutations.

The initial bank of polynucleotide sequences can consist of synthetic sequences that will be fragmented at step (a) or that can constitute the fragments of step (a).

According to a preferred embodiment of the process of the invention, step (a) consists in subjecting the polynucleotide sequences of the bank to hydrolysis by the action of one or several restriction enzymes.

To increase the degree of recombination produced by the process of the invention, it suffices to increase the number of restriction fragments by using restriction enzymes having a large number of cutting sites on the polynucleotide sequences of the bank, or by combining several restriction enzymes. If a thermostable and thermoactive ligase is used, the size of the

smallest fragment thus produced will advantageously be greater than or equal to 40 b or 40 pb, so as to maintain a hybridization temperature compatible with that of ligation step (d), which is generally on the order of 65°C.

Step (a) can further be done by producing a fragment bank by random enzymatic or mechanical treatment. In particular, step (a) can consist of a random treatment with DNAase I of a bank of polynucleotide sequences. If random enzymatic or mechanical fragmentation is used at step (a), this embodiment of the process of the invention is characterized in that it makes it possible to use fragments produced by this treatment as matrices for one another, for hybridization during step (c) or during the RLR reaction of steps (c) and (d) simultaneously.

Step (b) can be performed by combining at least two banks of distinct fragments produced separately in step (a) starting from the same initial bank by different treatments, such as, e.g., with different restriction enzymes. If such banks are used, the fragments obtained at step (a) are used as matrices for one another, for hybridization during step (c) or during the RLR reaction of steps (c) and (d) simultaneously.

The fragments of step (a) of the process of the invention can also be produced by amplification reactions (such as PCR) performed on the polynucleotide sequences of the bank. Two solutions in particular can be envisaged. In a first case, the initiated oligonucleotides can be designed so as to produce fragments whose ends are adjacent all along the assembly sequence. In a second case, the initiated oligonucleotides are designed so as to produce fragments having sequences in common, these fragments being able to act as an assembly matrix for one another at step (b) or at step (c).

The recombination efficiency of the process of the invention depends on the number of fragments produced per polynucleotide sequences at step (a). As

a result, the process of the invention will use polynucleotide sequences that have been fragmented into n fragments, n advantageously being greater than or equal to 3.

The assembly matrix of step (b) or (c) is, for example, a polynucleotide sequence produced from the initial bank or a sequence contained in said bank, single- or double-stranded. If the assembly matrix (matrices) is (are) incorporated directly at step (c) of the invention, this matrix must be in the single-stranded form.

According to a variant of the process of the invention, the assembly matrices of step (b) or (c) consist of single- or double-stranded oligonucleotides.

According to a particular form of the embodiment of the process of the invention, single- or double-stranded oligonucleotides of variable length are added at step (b) or (c) in addition to the matrix. These oligonucleotides are designed to be able to substitute for some of the fragments at step (c), in fact, their sequence is such that:

- if they are perfectly homologous with the sequence of the fragment they are replacing, they promote certain combinations, or
- if they are partially heterologous with the sequence of the fragment they are replacing, they introduce one or more supplementary, direct mutations.

Heterologous sequences are understood as two sequences whose base composition differs by at least one base.

Before step (e) of the invention, it is possible to separate the recombinant polynucleotide sequences from the assembly matrix thanks to a marker present on the assembly matrix or on the recombinant polynucleotide sequences. It is in fact possible to mark each strand of the matrix

according to techniques known to one skilled in the art. For example, the marker of the assembly matrix can be a hapten and the recombinant polynucleotide sequences are separated from the assembly matrix by techniques known to one skilled in the art, such as, for example, an antihapten antibody fixed on a carrier or a biotin-streptavidin reaction, if the hapten is a biotin marker.

Other techniques can be used to separate the recombinant polynucleotide sequences from the assembly matrix. The assembly matrix can also be prepared specifically so as to facilitate its elimination at the end of the process of the invention. It can thus be synthesized by PCR amplification using methylated dATP, which makes it possible to degrade it by the restriction endonuclease Dpn I. In this case, the recombinant polynucleotide sequences must not contain methylated dATP. The matrix can also have been prepared by PCR amplification by using some dUTP, which makes it possible to degrade it by treatment with a uracil-DNA-glycosylase. Conversely, it is possible to protect the recombinant polynucleotide sequences by amplifying them by selective PCR with oligonucleotides carrying phosphorothioated groups at 5'. A treatment with an exonuclease thus makes it possible specifically to degrade the assembly matrix.

The process of the invention can comprise, before the optional cloning at step (e), a step to amplify the recombinant polynucleotide sequences. Any amplification technique is acceptable, notably PCR amplification. One of the simplest consists in performing a PCR, which makes it possible to amplify specifically the recombinant polynucleotide sequences thanks to initiators that cannot be hybridized except at the ends of recombined sequences. The PCR products are then cloned, to be characterized, and the polynucleotide sequences with advantageous characteristics compared to corresponding

characteristics of reference sequences are selected.

The object of the invention is to produce polynucleotide sequences able to offer advantageous characteristics compared to corresponding characteristics of reference sequences. The recombinant polynucleotide sequences obtained at step (d) and optionally cloned are screened by any appropriate means to select the recombinant polynucleotide sequences or clones having advantageous characteristics compared with corresponding characteristics of reference sequences. Advantageous characteristics, for example, are understood to be thermostability of an enzyme or its ability to be able to function under pH or temperature conditions or saline concentrations better adapted to an enzymatic process than the reference proteins usually used for said process. By way of example of such a process, an industrial process of desizing textile fibers or bleaching of paper pulps or the production of flavors in the dairy industry, processes of biocatalysis to enzymatically synthesize new therapeutic molecules, etc., can be mentioned.

According to an advantageous embodiment of the process of the invention, the polynucleotide sequence bank can thus result from a screening having made it possible to select by any appropriate means the polynucleotide sequences offering advantageous characteristics compared to the reference sequences. The sequences thus selected constitute a restricted bank.

But it is also possible to start from a nonrestricted bank so as to maintain the representative nature of the characteristics the bank contains.

The sequences coding for the protein(s) having one or more advantageous characteristics compared to reference proteins are thus selected by *in vivo* or *in vitro* screens and can be used to form a new bank for an optional repeat of the process of the invention. An advantageous embodiment of the process

of the invention thus consists in using, as a bank, several polynucleotide sequences selected after a first running of the process of the invention, optionally mixed with other polynucleotide sequences. Among the screening techniques that can be applied to each clone of step (e), screening techniques by *in vitro* expression using notably *in vitro* transcription of recombinant polynucleotide sequences, then *in vitro* translation of the mRNAs obtained offer the advantage of eliminating cellular physiological problems and all the drawbacks connected with *in vivo* expression cloning. Further, this type of screening is easily automated, which makes it possible to screen a high number of recombinant polynucleotide sequences.

The invention also relates to a recombinant polynucleotide sequence obtained by a process according to the invention, as well as a vector containing such a recombinant polynucleotide sequence, a cellular host transformed by a recombinant polynucleotide sequence or a vector of the invention, as well as a protein coded by this recombinant polynucleotide sequence. The invention also comprises the corresponding banks of recombinant polynucleotide sequences, vectors, cellular hosts or proteins.

Example:

The object of this example is to produce recombinant polynucleotide sequences from the resistance gene to kanamycin by using oriented ligation of single-stranded fragments.

First, the resistance gene to kanamycin (1 Kb) of pACYC184 is cloned in the polylinker of M13mp18 so that the single-stranded phagemide contains the noncoding strand of the gene.

In parallel, this gene is amplified by PCR mutagenesis (error prone PCR) with two initiators that are complementary to vector sequence M13mp18 on each side of the gene sequence. The initiator for the noncoding strand is

phosphorylated while the initiator for the coding strand is not. The product of the PCR mutagenesis is digested by the lambda exonuclease, which produces a bank of coding strands for mutants of the resistance gene to kanamycin.

This bank of single-stranded molecules is digested by a mixture of restriction enzymes, notably Hae III, Hinf I and Taq I.

This bank of single-stranded fragments thus obtained is then hybridized with the single-stranded phagemide and ligated with a thermostable ligase. This step is repeated several times until the small fragments can no longer be observed during deposition on an agarose gel while the band corresponding to the single strand of the complete gene of resistance to kanamycin becomes a major component of the "smear" of single-stranded molecules visible on the gel.

The band corresponding to the size of the gene is then cut from the gel and purified. It is then hybridized with two complementary oligonucleotides (40 mer) of the M13mp18 sequences on each side of the gene and this partial duplex is digested by Eco RI and Sph I, then ligated in an M13 mp18 vector digested by the same enzymes.

The cells transformed with the ligation product are screened for increased resistance to kanamycin.

The cloning of single-stranded recombinant molecules can optionally be performed by PCR with two initiators of the complete gene and cloning of the double-stranded product of this amplification. To avoid undesirable mutations, this amplification will be performed with polymerase of the Pfu type and with a limited number of cycles.

The plasmids of the clones that are significantly more resistant to kanamycin than the initial stock are purified and used as matrices for PCR with the polymerase Pfu, under high fidelity conditions, with the

phosphorylated/nonphosphorylated initiator couple as previously defined. This produces the second generation of single-stranded fragments after a treatment with lambda exonuclease and fragmentation with restriction enzymes. The enzymes used for this step can comprise a different mixture (Bst NI, Taq I and Mnl I).

The recombination and selection steps are repeated several times until a substantial increase in resistance to kanamycin is obtained.